The Human Proteome Organization (HUPO) and Environmental Health

B. Alex Merrick

Proteomics Group, National Center for Toxicogenomics, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

The Human Proteome Organization, or HUPO, was formed to promote research and large-scale analysis of the human proteome. By consolidating national proteome organizations into an international body, HUPO will coordinate international initiatives, biological resources, protocols, standards and data for studying the human proteome. HUPO has identified five key areas to advance study of the human proteome, specifically in bioinformatics, new technologies, the plasma proteome, cell models, and a public antibody initiative. Consideration of three major issue areas may help develop HUPO's strategy for human proteome study. First is the need to distinguish the value of high throughput platforms from discovery platforms in proteomics. Second is the importance for international planning on integrating both transcriptome and proteome data and databases. Last is that effects of the environment from chemical, physical, and biological exposures alter the expression and structure of the proteome, which become manifest in long-term adverse health effects and disease. Environmental health research stands to greatly benefit from the shared resources, data, and vision of the HUPO organization as a valuable resource in exploiting knowledge of the human proteome toward improving public health. ***Key words:*** **environmental health, bioinformatics, human, proteome, proteomics, toxicogenomics, toxicology,** *Environ Health Perspect* **111:000–000 (2003). [Online 18 November 2002]**
doi:10.1289/txg.5918 available via ***http://dx.doi.org/***

The completion of the human genome is a major achievement of the 20th century. The 21st century challenge is to determine the function of the many newly discovered genes and how their gene products interact in pathways and systems to create the human body. An important approach in meeting this challenge in functional genomics is the use of large-scale analyses of the transcriptome and proteome. The human proteome is derived from the gene expression particular to a cell or tissue involving the dynamic and coordinated interaction of proteins in the body. "Mapping the human proteome" (Maher 2002) or describing the complex nature of protein structures, actions, and organizational hierarchies will be very unlike, and much more complex than, mapping the DNA sequence of the human genome. Multiple technologies and international cooperative strategies are being planned to meet the challenge of defining the human proteome and the subtle genetic variations reflected in protein polymorphisms that define each individual. This article summarizes proceedings of a new proteomics organization, comments on its goals and directions for the field of proteomics, and demonstrates why environmental health researchers have a vested interest in the agenda, cooperative studies, and shared resources that will emanate from this organization's activities.

The Human Proteome Organization, HUPO (2002), held an international meeting and workshop at the National Institutes of Health on 29 April 2002 to prioritize goals and standards for large-scale analysis of the human proteome. The mission of the organization is to *consolidate* national proteome organizations into the international body HUPO; to *engage* in scientific and educational activities that promote technologies pertaining to the human proteome and model organisms; and to *assist* in coordinating shared, public proteomic initiatives. The president of the organization is Samir E. Hanash at the University of Michigan. Currently, member countries are linked by three international HUPO divisions: North America, Europe, and Asia–Oceania, with countries from all three divisions participating at the workshop. The two major challenges for HUPO are to identify major opportunies first for international cooperation and second for joint initiatives between public and private sectors.

The HUPO meeting focused on developing specific agendas in five key areas (Figure 1) of human proteomics for immediate international development, chaired by recognized leaders in the field: "Bioinformatics," Rolf Apweiler (University of Heidelberg, Heidelberg, Germany; EMBL, European Bioinformatic Institute, Hinxton, Cambridge, UK); "New Technology," Richard Simpson (Ludwig Institute, Melbourne, Victoria, Australia); the "Plasma Proteome," Gilbert S. Omenn (University of Michigan, Ann Arbor, Michigan, USA); "Cell Models and Tissues," Ronald Taussig (University of Michigan), Cell Signaling Alliance and Pei Pei Ping (University of California, Los Angeles, California, USA); and the "Antibody Initiative," Mattihias Mann (Odense University, Odense, Denmark). A brief discussion of each area follows.

Bioinformatics

The bioinformatics group will define the proteomic data platforms such as 2D (two-dimensional) gels, protein arrays, mass spectrometry, and structural data into a defined infrastructure for data submission and annotation. A major bioinformatics issue is to determine a direction toward either a linked, interoperable consortium of small distributed proteomics databases or the alternative of a large, centralized database. Annotation standards need to be defined using a controlled vocabulary and data confidence measures. Because journals contain many raw and processed proteomic data for potential database incorporation, copyright and accessibility issues need to be resolved.

New Technology

The new technology group had several objectives that included determining lead technologies for best discovering protein interactions, quantifying proteins over a wide dynamic range, fractionating cellular and subcellular compartments to acceptable levels of purity, and identifying housekeeping genes for normalization. The group plans to establish web-based HUPO protocols and make available sets of protein standards that are platform-independent. The group was interested in whether high throughput technologies could be developed to define protein states such as posttranslational modifications, protein conformations, cellular localization, splice variants, covalent modifications, proteoloysis, and ligands. A goal was set to identify

Address correspondence to B.A. Merrick, NIEHS, D2-04, PO Box 12233, Research Triangle Park, NC 27709 USA. Telephone: (919) 541-1531 Fax: (919) 541-4704. E-mail: merrick@niehs.nih.gov

I thank M.D. Waters and E.K. Lobenhofer for a critical reading of the manuscript and recognize R.W. Tennant, J.K. Selkirk, and K.B. Tomer for creative discussion and critique. I also thank B. Mills of Image Associates, Raleigh, North Carolina, for the graphics artwork.

Received 31 July 2002; accepted 18 November 2002.